Service Assurance in Virtualized Data Centers

A foundation for cloud services and the Software-Defined Data Center (SDDC)

Introduction

IT organizations are increasing their adoption of virtualization technologies as a means of transforming their data centers to meet the goal of delivering business services to users with far greater efficiency and responsiveness than is possible in a traditional data center. A well-designed virtualized data center offers a number of proven benefits, including the consolidation of physical resources and the dynamic optimization of workloads across a multi-domain environment. It also enables the automation of administrative processes that results in either minimizing or eliminating the need for operator involvement in well-understood, standardized tasks.

However, in order to realize the full potential of a virtualized data center, organizations also need to transform their management capabilities to accommodate the complexities of an environment that is comprised of a mix of physical and virtual resources. In particular, management processes and the associated tool sets for key functions such as provisioning, configuration, and service assurance need to have capabilities that span the boundaries between physical and virtual resources as well as the boundaries between the various technology domains.

Managing service assurance in a virtualized environment can be very challenging. This is especially true in an environment in which an IT organization is working with a set of management tools and each tool is focused only on a single technology domain or only on one side of the physical/virtual boundary. Working with tools of that nature typically results in significantly limiting the visibility that IT organizations have into the operation of their virtualized data centers. This lack of detailed visibility makes it extremely difficult to determine in which technology domain the root cause of a problem resides and that in turn results in increasing the amount of time it takes to resolve problems. This lack of detailed visibility also makes it extremely difficult to understand the business impact of a problem and that negates the ability of the IT organization to effectively prioritize trouble tickets.

The goal of this white paper is to provide some insight into the challenges of managing service assurance in a virtualized data center and then to describe how the unique capabilities of EMC's Service Assurance Suite (SAS) can help data center operations groups meet these challenges and optimize the level of service that they deliver to end users,

Trends in the Virtualized Data Center (VDC)

As described in the 2012 Cloud Networking Report[1], the most significant factor in the evolution of data centers is the ongoing virtualization of multiple components of the IT infrastructure, including application servers, appliances, storage and networks. In addition to providing better resource utilization and lower CAPEX, virtualization creates a dynamic environment that can respond in real time to changes in user demand for application services. The current level of server virtualization provides the foundation for both public cloud *Infrastructure as a Service* (IaaS) offerings as well as private cloud computing environments. However, in order to deliver a

[1] http://www.webtorials.com/content/2012/12/2012-cloud-networking-report-1.html

truly elastic infrastructure, the rest of the data center must be virtualized as well. The ongoing virtualization of the infrastructure will enable the implementation of a Software-Defined Data Center (SDDC) in which all of the components of the infrastructure are virtualized and delivered as a service, and the control of this data center is entirely automated and orchestrated by software.

Just as server virtualization has delivered higher levels of agility and efficiency for compute resources, in a Software Defined Data Center the abstraction of network, security and storage resources will move higher levels of functionality from hardware to software to deliver an agile, elastic infrastructure. By taking logical abstractions of these data center resources, pooling these resources into managed entities and automating tasks across the data center, IT organizations can deliver a higher degree of scalability and efficiency across both enterprise and service provider data centers. This approach also centralizes the capture of telemetry across the virtual constructs to ensure that service levels within the virtual data center are being maintained.

As server virtualization has evolved to become a mainstream data center technology, the majority of servers in data centers are now virtualized and server consolidation ratios (i.e., the number of virtual machines (VMs) per physical server) are growing rapidly due in large part to the growth in the number of processor cores and I/O bandwidth per server CPU. With more servers being virtualized, IT organizations are making greater use of the live migration of VMs (e.g., with VMware's vMotion capability) as a means of automating and optimizing workloads and operational tasks in order to make the infrastructure more dynamic and efficient. This includes:

- Managing the loading of virtualized physical servers
- Expanding and contracting application capacity dynamically to meet fluctuations in demand
- Performing zero downtime scheduled maintenance and zero downtime repair of impaired, but still functional, servers
- Reducing power and cooling expenses during periods of light server loading

In the early phases of the adoption of server virtualization within data centers, it was somewhat common for IT organization to identify a number of applications as being too mission-critical or too resource-intensive to be supported on a virtualized infrastructure. However, as virtualization has matured and multi-core physical servers have emerged with greatly increased processing power and I/O capacity, all types of applications are currently being deployed on virtualized servers, storage and network resources. As this trend continues, multi-tiered enterprise applications, such as ERP, and real-time, latency-sensitive applications, such as VoIP, will increasingly be deployed on virtualized infrastructure. As the infrastructure becomes pervasively virtualized and essentially all enterprise applications are dependent on virtualized resources, IT organizations must place much greater emphasis on virtualization-aware service assurance solutions.

Operational Challenges for Service Assurance in a VDC

As described in the 2012 Application & Service Delivery Handbook[2], providing an assured level of service for critical enterprise applications has always been something of a challenge for IT organizations, and this challenge often results in either the lack of SLAs or of SLAs that are less robust than end users want. While the issue can be partially addressed with a highly redundant, highly available infrastructure design, the ability to provide predictable levels of service is ultimately dependent on the following operational management capabilities.

Automated, Dynamic Discovery of Relationships

In a traditional data center, a fundamental requirement for network operations groups is the automated discovery of all the elements of the infrastructure as well as the discovery of the complete topology of how the elements are connected. In order to manage service assurance in a VDC, network operations groups need the automated ability to understand the dynamic relationships between the elements of the VDC that need to be managed.

Ideally, discovery can be accomplished primarily by using standard protocols, such as SNMP, with minimal if any reliance on special agents or proprietary APIs or protocols. While topology discovery can be problematic in a large, diverse physical environment, understanding the dynamically changing relationships between the elements of a VDC is significantly more challenging. This follows because the layers of abstraction that are associated with the virtualization can impair visibility into the dynamically changing, end-to-end physical/virtual, server/storage/network infrastructure upon which a virtualized application or service relies.

The challenges associated with discovering topology and understanding relationships in a VDC can only be met by implementing a discovery engine that is capable of learning the full cross-domain topology of the data center, including all the physical and virtual elements of the infrastructure.

Root Cause Analysis

Data center operations must be able to quickly identify and isolate the cause of any problem that result in either impaired application availability or performance. This means that IT organizations must have the capability to minimize both the Mean Time to Identify (MTTI) and the Mean Time to Resolve (MTTR) any problem. Even in purely physical data centers root cause analysis can be a challenge if the analysis engine doesn't support the full gamut of devices in the infrastructure or where the analysis engine is based on rules which the IT organization must customize to accommodate each specific network topology. Rule customization is time consuming, error prone, and may have to be performed each time a change is made to the infrastructure.

[2] http://www.webtorials.com/content/2012/08/2012-application-service-delivery-handbook-2.html

In a virtualized data center, root cause analysis is made more complex because, as previously mentioned, the relationships between elements are typically obscured by the abstraction that is associated with the virtualization of the infrastructure. As a result, when service performance is degraded or availability is impaired, it is very difficult for an IT organization to sort through a flood of symptomatic alerts to determine the root cause of the problem—or even in which technology domain the root cause can be found

The challenges associated with performing root cause analysis in a VDC can be met with an advanced root cause analysis engine that leverages cross-domain, physical/virtual topology discovery and which can quickly isolate the cause of faults or degraded functionality and/or performance in very large and complex network topologies. Because of the dynamic nature of virtualized environments, the correlation engine should not rely on rules that need to be modified by operators in order to accommodate specific topologies.

Impact Analysis

When a problem occurs in the infrastructure, the IT operations group must address two questions: what is the root cause of the problem (discussed above) and what priority should be placed in addressing the issue. The answer to the second question generally depends on an analysis of the services and the users that are impacted. Similar to the previously discussed challenges that are associated with root cause analysis, the relationships between the users, the application services and the physical infrastructure can also be obscured by virtualization, making it difficult to assess the severity of the problem.

To respond to this challenge, IT organizations need to implement an impact analysis correlation engine that can leverage the virtualization-aware discovery engine in order to map bilaterally between the services being delivered and the supporting infrastructure**.**

Automated Change and Configuration Management

Historically change and configuration management have been manual processes and as a result they have been time consuming, resource intensive and expensive. Because those processes are manual it makes it difficult for an IT organization to determine whether or not their infrastructure is in compliance with regulatory requirements and industry best practices. The manual nature of change and configuration management also means that those processes tend to be error prone to the point that in the current environment it is widely accepted that improper configuration or change management is the cause of the vast majority of service-affecting problems. In some instances, these service-affecting problems are outages of some type, such as a network switch being unavailable. While these types of service-affecting problems can negatively impact business critical processes, they are relatively easy to identify and resolve.

In many other instances, however, these service-affecting problems don’t result in an outage, but in degraded performance such as would occur if because of an improper change in the configuration of a switch, the performance of that switch, or one or more ports on that switch was significantly reduced. These types of service-affecting problems can also have a negative

impact on business critical processes, but unfortunately identifying and resolving these types of problems is very difficult. That difficulty stems from the fact that in most situations IT organizations don't have a good understanding of the relationships between the wide range of elements that comprises their infrastructure.

On a going forward basis, the negative impact of improper configuration or change management will increase. That follows in part because as previously mentioned, in a virtualized data center the relationships between elements are typically obscured by the abstraction that is associated with the virtualization of the infrastructure. Eliminating this negative impact requires an automated approach to configuration and change management that is applicable from network design through production, and which applies in both a physical and a virtual environment.

Performance Management

A critical component of service assurance in any type of IT environment is the ability to manage the service levels that the IT organization is expected to provide, whether those service levels have been documented in an explicit SLA or are part of the implicit expectations of the company's business unit managers. In order to support effective service level management, a performance management tool must have visibility into networks, systems, storage and applications. The tool must provide both historical and real-time visibility and it must also be able to project future performance.

An effective performance management tool must also provide sophisticated analytics and reporting. In particular, the tool must be able to identify potential problems before they impact users and must be able to ensure that the appropriate personnel within the organization informed of the potential problem.

EMC Service Assurance Suite (SAS)

EMC SAS is an integrated service assurance solution, comprised of EMC Smarts and Watch4net products, and has been designed expressly for cross-domain topology discovery, root cause analysis, performance management, and impact analysis in highly virtualized data center environments. EMC SAS uses SNMP polling to monitor the majority of the physical components of the data center network. For virtual elements, EMC SAS relies on two-way communications with hypervisor management systems via native APIs; e.g., the VMware API for vCenter. In addition, the Cisco Discovery Protocol (CDP) or the Link Layer Discovery Protocol (LLDP) are used where needed to augment SNMP for virtual switch and edge switch discovery.

EMC SAS has a unique combination of capabilities that enable data center operations personnel to support very high levels of service assurance. These capabilities include:

Topology Discovery and Monitoring

This capability includes the automated discovery and monitoring of all of the physical and virtual elements on the data center network. At the virtualized edge of network, the SAS discovery

engine provides full visibility of application VMs, virtual appliances, hypervisors/servers, vNICs, vSwitches, PNICs and access switches as well as the relationships amongst these entities. Automated continuous discovery keeps track of both topology changes and new elements that are added to the network. SAS provides support for both VMware and Microsoft hypervisors.

As part of the discovery process, SAS has the unique capability to discover all of the storage resources used by VMware's vCenter; e.g., iSCSI SAN, Fibre Channel SAN and NAS. For Ethernet-attached storage, SAS provides full logical and physical visibility for iSCSI and NAS I/O via the ESX VMKernel Port. For Fibre Channel (FC) attached storage, SAS provides full visibility of everything on the logical FC SCSI path, including the VM, ESX hypervisor, HBAs, SAN fabric, Data Store and SCSI LUN. SAS integration between vSphere 5.x. and the EMC Storage Resource Management (SRM) Suite can extend FC visibility beyond the logical path to the physical path, including FC switches, ports and cables. The SRM API also provides the capability to monitor storage performance and generate threshold alerts that enable the proactive identification of storage issues that can affect service levels. SAS also discovers multi-path connections among data center elements, including Ethernet Link Aggregation Groups (LAGs), EMC PowerPath storage connectivity for physical servers, and VMware's Native Multi-pathing (NMP) or EMC's PowerPath/VE for virtualized servers.

The integration with vCenter for monitoring both VMotion and Storage VMotion allows dynamic topology updates for the location of VMs and the location of their data stores. SAS discovers both Distributed Resource Scheduler (DRS) server clusters and Storage DRS (SDRS) storage clusters. If vMotion is used, SAS can provide the operations personnel with insight into why a VM was moved; i.e. automated movement via DRS or operator initiated.

vSphere Distributed Power Management (DPM) continuously optimizes power consumption in the data center. When VMs in a DRS cluster need fewer compute resources, such as during nights and weekends, DPM consolidates workloads onto fewer servers and powers off the rest in order to reduce power consumption. When VM resource requirements increase, DPM brings the powered-down hosts back online in order to ensure that service levels are met. With Storage VMotion, SAS's topology map is automatically updated when SDRS moves a VM's data store within the storage cluster, whether that movement is based on disk space utilization, disk latency or LUN performance.

SAS also performs Cisco-certified discovery and monitoring of Unified Computing System (UCS) blade server internals; e.g., Fabric Interconnect and fabric extenders, virtualized NICs, blade/host relationships as well as chassis entities including fans, power supplies, temperature and voltage sensors. For physical rack servers from a wide variety of vendors, SAS monitors native server agents via the SNMP Host Resources MIB or via Windows Management Instrumentation (WMI).

SAS Application Process Monitoring leverages SNMP to provide operational visibility into the status of applications and processes on physical and virtual servers. It generates alerts when a process is missing, stopped or exceeds process and/or application thresholds. It cross-correlates these application alerts with the state of infrastructure, as described in the next section of this white paper. Default templates are available for automated discovery and monitoring of key

processes for ESX 6.x, vCenter 4.x and 5.x, and EMC ATMOS servers. Operations personnel can modify these templates and can also configure SAS Application Process Monitoring to group processes that comprise a business service.

Root Cause Analysis

The SAS root cause analysis is based on behavioral models of physical and virtual network elements in the server/storage/network domains in conjunction with Codebook Correlation Technology™ (CCT). CCT is a patented technology that diagnoses problems in real-time by matching symptoms to problem signatures. CCT diagnoses problems by combining generic behavioral models of various classes of data center elements with the discovered topology map of the infrastructure. For each possible problem, the behavioral models include both the local symptoms (i.e., those displayed within the device itself) and propagated symptoms (i.e., those displayed within other devices in the surrounding topology).

CCT uses a special algorithm to automatically compute a complete set of problem signatures based on local symptoms of the problem and all the additional symptoms that are propagated throughout the discovered topology. The resulting problem/signature mappings are then stored in the CCT codebook.

When a problem occurs, CCT uses the codebook to determine the problem whose signature matches the set of symptoms being reported. This is declared to be the root cause of the problem. When an exact codebook match is not made, CCT identifies the root cause as the problem whose signature has the highest likelihood of matching the set of symptoms.

Compared to conventional RCA correlation engines based on rules or downstream event de-duplication, CCT offers a number of significant advantages:

- **Out-of-the-box Deployment:** Because the correlation logic is automatically computed from reusable object-oriented behavior models in conjunction with the automatically discovered network topology, CCT RCA works out-of-the-box and it also adapts to the network without any development or maintenance as the topology evolves.

.

- **Scalability:** Since the processing of each event involves relatively simple codebook lookups, CCT correlation logic can handle many hundreds of events per second, separating root causes from symptoms in very large, complex data center infrastructures.

- **Accuracy:** Because CCT looks for the closest match between observed events and problem signatures, it can determine the root cause even from an incomplete set of symptoms, e.g., when some event notifications are delayed or lost.

Risk Analysis

In addition to gathering symptoms of actual problems, SAS infrastructure monitoring can identify a wide variety of risk conditions that can provide an early warning of a possible impending problem. These risk conditions include:

- A partial failure within redundant subsystems; e.g., power supplies, fans, or path groups such as Ethernet LAGs and PowerPath/VE. SAS allows operators to set thresholds that will trigger risk condition alerts.

- Application Process Monitoring thresholds exceeded.

- Host and Blade Server environmental variables beyond threshold.

Impact Analysis

SAS Business Impact Manager performs bi-directional mapping between business processes/services and the infrastructure upon which these services rely. As a result, problems or risk conditions with applications or within the infrastructure can be mapped to the business services and the users that are impacted. The SAS Business Impact Manager can also calculate an aggregate severity measure for each problem based on user-assigned criticality metrics that define the business importance of each of the services and users that are affected.

The severity of the issue can be used to prioritize allocation of resources to the resolution of any problems that arise throughout the infrastructure. In many cases a risk condition with high potential for severe business impact can be resolved before services or users are disrupted.

As a result of the integration of the SAS topology discovery, RCA, and impact analysis functions a single console screen can provide a global view of symptoms, root cause, impacted services and users/customers, as shown in Figure 1

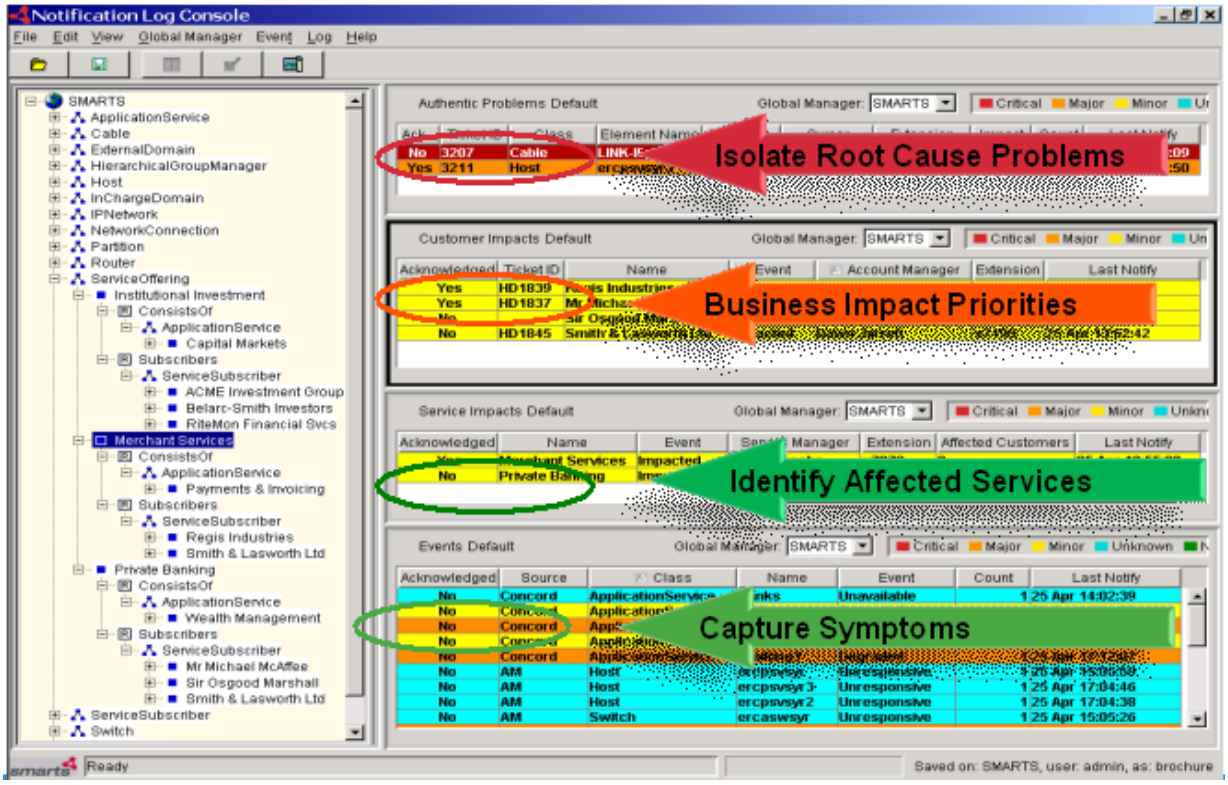

Figure1: Global View of Symptoms, Root Cause, and Impact Analysis Results

EMC SAS Network Configuration Management

The EMC SAS provides a model-based, multi-vendor, automated compliance, change and configuration management solution. The tool automates support for all of the stages of the network infrastructure lifecycle by integrating the required design, change and compliance requirements. At the design stage, the EMC SAS provides a virtual design workspace that enables IT organizations to design new virtual networks that are based on their existing designs. It also utilizes intelligent automation capabilities to allow IT organizations to quickly create error-free, large-scale designs.

During the production stage of the network infrastructure lifecycle the EMC SAS uses golden configurations to create templates that IT organizations can use for new device deployments. This capability, combined with ITIL-compliant change processes and workflow approvals, eliminates virtually all standard change errors. The tool can also be adapted to specific environments and can be integrated with common workflow, trouble-ticketing and help-desk solutions.

IT organizations have to deal with compliance requirements throughout the network infrastructure lifecycle. The EMC SAS enables IT organizations to respond to those requirements in part by enabling IT organizations to enforce policies over the entire IT infrastructure, not just one network, one site or one subnet. The tool also allows IT organizations

to report on the historical compliance of the infrastructure using the configurations and policies that were in place on a given date.

As shown in Figure 2, the Web-based interface provides easy access to the critical compliance, change and configuration management information that IT organizations require.

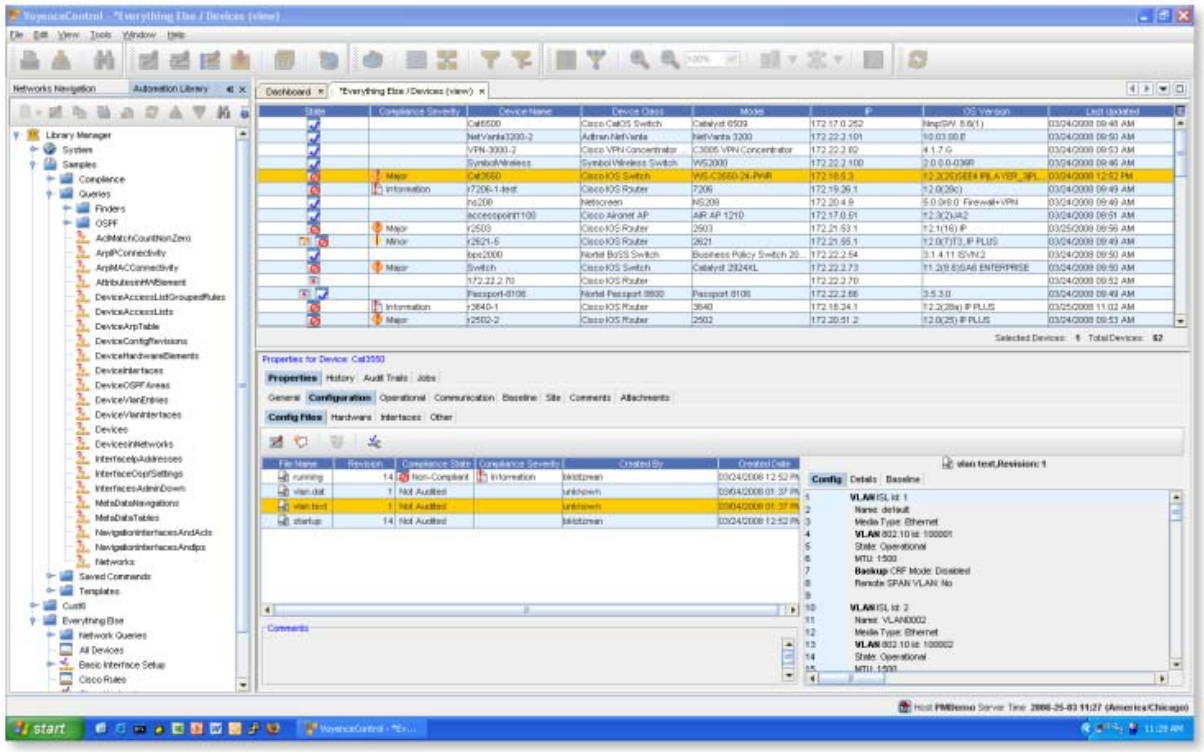

Figure 2: Web-based Access to Compliance, Change and Configuration Management Information

Performance Management

The EMC Watch4net is a carrier-class performance management solution that provides IT organizations with visibility into the historical, real-time and projected end-to-end performance of their network, compute and storage resources. The tool provides IT organizations with comprehensive Web dashboards and targeted reports that enables business and technical users to see in real-time the status of the IT infrastructure and the business services that are supported by that infrastructure.

The views provided by EMC Watch4net are categorized by topics such as inventory, performance, security and trending. Because the tool combines technology and business metrics, Watch4net enables organizations to manage their explicit and implicit SLAs. As performance data is collected, Watch4net continually analyzes in real-time thousands of metrics across the entire IT infrastructure. Any suspicious activity is included in a global report entitled “situations to watch”. Alerts are sent as soon as a potential problem is detected using either fixed or adaptive thresholds.

Summary & Conclusion

Along with its many benefits, virtualization of the data center adds significantly to the complexity of the infrastructure, which results in a number of challenges for operations personnel. Perhaps the greatest challenge is assuring acceptable levels of availability and performance for critical business services, which heavily depend on the efficiency of tools and procedures to isolate and resolve problems that arise in the infrastructure. Minimizing MTTI and MTTR is especially challenging where operations personnel must rely on separate, non-integrated service assurance tools to monitor physical and virtual resources or where separate tools are used in each technology domain.

EMC SAS is a single integrated solution that includes all the key functions required for service assurance across both physical and virtual entities within the three major technology domains: servers, storage and networks. SAS's combination of automated topology discovery, advanced CCT root cause analysis, application process monitoring, risk analysis, impact analysis, change and configuration management as well as performance management allow IT operations groups to quickly identify, prioritize, and resolve any problems that arise in very large complex virtualized data centers. Application Process Monitoring and Risk Analysis complement the extensive monitoring capabilities to enable proactive circumvention of many nascent problems before they reach the point of disrupting services and users.